

Keterkaitan Antara Variabel dan Prediksi Nilai Indeks Pembangunan Manusia (IPM) di Indonesia Menggunakan Regresi Linier

Nanda Cahaya Putra¹, Ikram Yunizar², Ndaru Aji Laksono³, Nurul Chamidah⁴

Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jakarta
email: ikramyunizar@gmail.com
Jl. Rs. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia

Abstrak

Penelitian ini bertujuan untuk membuat pemodelan yang dapat digunakan untuk memprediksi nilai IPM serta variabel apa yang mempengaruhi nilai dari Indeks Pembangunan Manusia (IPM). Dalam penelitian ini menguji variabel terikat dengan beberapa variabel bebas yang memiliki keterkaitan dengan Indeks Pembangunan Manusia (IPM) dengan menggunakan dua buah skenario. *Dataset* yang dipakai dalam penelitian ini merupakan beberapa data yang memiliki keterkaitan terhadap Indeks Pembangunan Manusia (IPM) dengan periode 2010 – 2017. Penelitian menggunakan metode regresi linier untuk menentukan model. Penelitian menggunakan dua skenario dalam membagi data menjadi data uji dan data latih yang memiliki jumlah *record* data uji dan data latih yang berbeda. Hasil dari penelitian ini menunjukkan bahwa pemodelan dengan metode regresi dapat diterapkan pada data yang dipakai dan mendapatkan hasil yang baik. Pemodelan tersebut kemudian di evaluasi menggunakan *R-Square* dan *Root Mean Square Error* (RMSE) untuk mengetahui seberapa besar pengaruh variabel terkait terhadap Indeks Pembangunan Manusia (IPM).

Kata kunci: IPM, regresi linier, prediksi, keterkaitan data

1. PENDAHULUAN

Indeks Pembangunan Manusia (IPM) adalah pengukuran perbandingan dari harapan hidup, melek huruf, pendidikan dan standar hidup untuk semua negara diseluruh dunia (UNDP. 1990). Penelitian terhadap IPM dan regresi linier telah banyak dilakukan sebelumnya, salah satunya adalah penelitian Amrin menggunakan metode regresi linier berganda untuk memprediksi tingkat inflasi bulanan pada masa yang akan datang. Data dilakukan pra-proses menggunakan *Cross-Standard Industry for Data Mining* (CRISP-DM). Dilakukan uji regresi linier berganda untuk mendapat nilai prediksi dari model. Prediksi tersebut dievaluasi dan mendapatkan *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE) dan *Mean Absolute Deviation* (MAD) sebagai penentu untuk memprediksi inflasi bulanan yang akan terjadi pada masa yang akan datang. Darsyah (2017) melakukan klasifikasi beberapa variabel terkait dengan IPM yang ada di Jawa Tengah menggunakan algoritma *k-Nearest Neighbor* (k-NN). *Dataset* yang telah dikumpulkan dibagi menjadi data *training* dan *dataset stratified 10 fold-cross validation* dengan tiga buah pendekatan menggunakan algoritma *k-Nearest Neighbor* menghasilkan nilai rata-rata IPM di Provinsi Jawa Tengah pada tahun 2016. Penelitian Putra dan Vita (2015) terjadi kasus *multicollinearity* pada faktor-faktor yang mempengaruhi IPM di Jawa Timur. Regresi logistik ridge dipilih untuk mengatasi kasus tersebut dengan cara memprediksi model yang terbaik menggunakan *ridge parameters*. Dari penelitian tersebut hasilnya terdapat 5 kabupaten/kota tergolong IPM menengah bawah., sedangkan 33 kabupaten/kota lainnya tergolong IPM menengah atas. Sari dan Priati (2016) memodelkan lima belas variabel yang memiliki

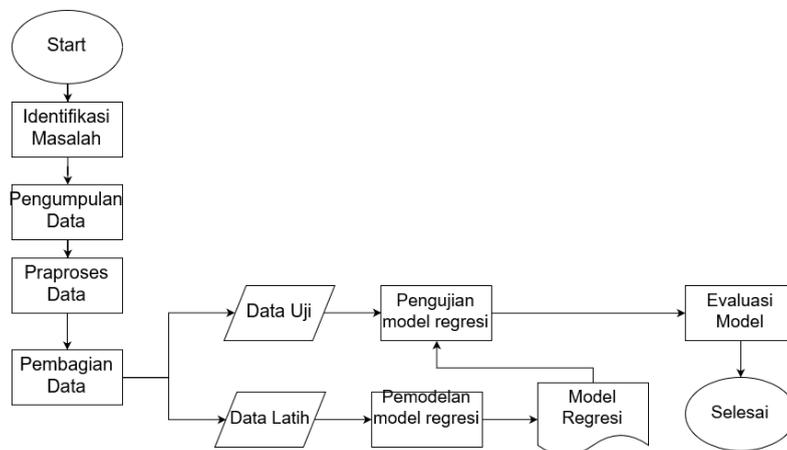
keterkaitan dengan tingkat IPM yang ada di Jawa Barat. Metode *Dynamic Bayesian Network* (DBN) digunakan dalam memodelkan variabel terkait. *Dataset* yang diperoleh kemudian diolah melalui beberapa tahap, yaitu tahap praproses, dan membagi *dataset* menjadi dua buah skenario. Evaluasi dilakukan terhadap kedua skenario tersebut untuk memprediksi kategori tingkat IPM. Hasil penelitian ini memberikan informasi tentang faktor yang perlu diperhatikan untuk meningkatkan IPM yang ada di Jawa Barat.

Dari beberapa jurnal diatas, belum ada penelitian IPM yang mencakup ruang lingkup negara Indonesia. Penelitian ini dilakukan untuk memberi informasi tentang nilai faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia di Indonesia secara keseluruhan. Beberapa penelitian hanya mencakup beberapa provinsi yang ada di Indonesia. Metode regresi linier dipilih karena menurut Amrin (2006) metode regresi linier dapat melakukan generalisasi dan ekstraksi dari pola data tertentu, mampu mengakuisisi pengetahuan walau, tidak ada kepastian, dan mampu melakukan perhitungan secara paralel sehingga proses lebih singkat.

Pada penelitian ini akan dilakukan pemodelan dan prediksi pada *dataset* yang telah dikumpulkan dengan beberapa tahapan. Tahapan tersebut diantaranya akan melakukan praproses data menggunakan *mean* sebagai metode untuk mengisi *missing value* pada *dataset* yang dikumpulkan. Tahap selanjutnya, dilakukan pemodelan dan prediksi dengan menggunakan metode regresi linier berganda. Pada penelitian diatas, digunakan berbagai macam algoritma dan metode dalam menentukan model pada penelitian. Salah satunya regresi berganda, pemodelan dan prediksi menggunakan metode tersebut mendapatkan hasil yang baik, maka metode regresi linier berganda dipilih dalam penelitian ini. Penelitian ini diharapkan dapat memberi informasi berupa faktor-faktor yang harus diperhatikan dalam upaya meningkatkan IPM di Indonesia dan memberikan dasar acuan untuk penelitian lebih lanjut tentang IPM di Indonesia.

2. METODOLOGI PENELITIAN

Penelitian ini memiliki beberapa tahapan yang dilalui untuk mendapatkan hasil yang diharapkan, tahapan-tahapan yang harus dilalui dapat dilihat dalam *flowchart* pada gambar 1.



Gambar 1 : *Flowchart* tahapan penelitian

1.1 Tahap Identifikasi Masalah

Dalam penelitian ini yang menjadi masalah adalah faktor apa saja yang dapat mempengaruhi nilai IPM seluruh provinsi yang ada di Indonesia dan memprediksi nilai IPM berdasarkan faktor tersebut.

1.2 Tahap Pengumpulan Data

Data yang tersaji dalam penelitian ini diperoleh dari *website* Badan Pusat Statistik (www.bps.go.id). Data yang telah terintegrasi memiliki 272 *record* yang berisi nilai masing-masing

variabel dari 34 provinsi yang ada di Indonesia dari tahun 2010 -- 2017. Variabel-variabel yang terdapat dalam penelitian ini sebagai berikut.

1. Harapan hidup;
2. lama sekolah;
3. kesehatan;
4. kemiskinan;
5. Indeks Pembangunan Manusia.

1.2.1 Harapan Hidup

Harapan hidup adalah rata-rata tahun yang dijalani oleh seseorang yang berhasil mencapai umur tertentu, pada tahun tertentu, dalam situasi mortalitas yang berlaku di lingkungan masyarakatnya.

1.2.2 Lama Sekolah

Angka lama sekolah adalah jumlah tahun belajar penduduk berusia 15 tahun keatas yang telah diselesaikan dalam pendidikan formal. Angka lama sekolah dapat menunjukkan jenjang pendidikan yang pernah atau sedang diduduki oleh seseorang.

1.2.3 Kesehatan

Nilai-nilai dari variabel ini menunjukkan nilai perbandingan jumlah orang sakit, baik secara fisik maupun kejiwaan pada suatu provinsi dengan jumlah seluruh penduduk yang ada di provinsi tersebut pada tahun tertentu.

1.2.4 Kemiskinan

Variabel ini menunjukkan nilai perbandingan dari jumlah penduduk yang tidak dapat memenuhi kebutuhan dasar dengan jumlah penduduk pada provinsi dan tahun tertentu.

1.2.5 Indeks Pembangunan Manusia

Variabel yang berisi nilai Indeks Pembangunan Manusia dan menjadi variabel terikat dalam penelitian ini. Nilai-nilai pada variabel ini didapatkan dari nilai IPM yang ada di provinsi Indonesia pada rentang tahun 2010 – 2017. Nilai IPM dapat menunjukkan kualitas sumber daya manusia di masing-masing provinsi di Indonesia.

1.3 Tahap Praproses Data

Ada beberapa tahapan yang dilalui dalam tahap praproses data, antara lain integrasi data, pengisian *missing value*, dan normalisasi. Pada tahap integrasi data, data yang diperlukan untuk mendapatkan model penelitian dikumpulkan dan disatukan. Data perlu disatukan karena data diambil dari beberapa *dataset* berbeda. Setelah data disatukan, didapatkan variabel-variabel yang akan digunakan untuk pemodelan.

Setelah data diintegrasikan, *record* pada data akan diperiksa apakah ada nilai-nilai kosong atau *missing value* yang dapat mengubah hasil penelitian secara keseluruhan. Dari hasil pemeriksaan, didapatkan beberapa *missing value*. *Missing value* ada di dalam *record* yang didapatkan dari provinsi Kalimantan Utara dari tahun 2010 -- 2012. Provinsi Kalimantan Utara tidak memiliki nilai data dalam beberapa variabel karena provinsi ini baru dimekarkan pada tahun 2012. Untuk mengisi nilai dari provinsi ini, data yang kosong diisi dengan nilai *mean* dari keseluruhan data yang ada. Rumus untuk menghitung nilai *mean* menurut Bardja (2017) adalah.

$$\bar{x} = \frac{\sum x}{n} \quad (1)$$

Keterangan :

\bar{x} = nilai *mean*

x = nilai masing-masing data

n = jumlah data

Setelah *missing value* telah diisi, maka tahapan selanjutnya adalah analisis tentang *range* masing-masing variabel. Seluruh variabel memiliki *range* angka antara 0 sampai 100, sehingga variabel yang ada tidak perlu dinormalisasi. Dari data yang tersaji diketahui nilai *mean* variabel IPM sebesar 67.72, nilai tersebut akan digunakan untuk mencari besarnya kesalahan.

1.4 Tahap Pembagian Data

Setelah tahap pra proses selesai, data yang dihasilkan akan diolah melalui dua skenario pembagian data sebagai berikut.

1. Skenario 1 : Membagi data menjadi data latih yaitu *record* yang diambil pada tahun 2010 – 2016 dan data uji diambil dari *record* tahun 2017
2. Skenario 2 : Membagi data dengan metode *k-fold cross validation* dengan menentukan nilai k adalah 3

Pembagian skenario ini dilakukan untuk memeriksa apakah model yang dihasilkan metode regresi linier dapat memberikan hasil yang konsisten pada setiap pembagian data. Kedua skenario akan menghasilkan jumlah *record* data latih dan data uji yang berbeda yang akan dijadikan acuan untuk pengujian model.

1.5 Tahap Pemodelan

Pada tahap ini, data yang telah dibagi menjadi data latih dan data uji akan diolah dengan metode regresi linier untuk mendapatkan model yang dapat digunakan untuk memprediksi hasil dan menentukan keterkaitan antara variabel yang ada. Syilfi dan Diah Safitri (2012) merumuskan bentuk dasar dari regresi linier antara lain.

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n \quad (2)$$

Keterangan :

Y = variabel terikat

α = konstanta

β = koefisien variabel

x = variabel bebas

Setelah model data telah didapatkan melalui metode regresi linier, maka model yang ada akan diuji ke dalam data uji untuk memprediksi nilai Indeks Pembangunan Manusia dari model yang telah didapatkan dan dievaluasi.

1.6 Tahap Pengujian

Ketika model sudah didapatkan dari kedua skenario pembagian data, model akan diuji dengan memprediksi nilai IPM masing-masing pada data uji. Pengujian model dilakukan dengan menggunakan *scatter plot* yang menampilkan berbagai informasi tentang kelayakan model pada data. *Scatter plot fitted values vs residuals* akan menampilkan informasi tentang selisih angka antara nilai prediksi dan nilai observasi pada tiap record dan menggambarkan sebaran data secara keseluruhan. Menurut Sinambela *et al.* (2014), *residual* dapat dihitung dengan rumus.

$$e_i = y_i - \hat{y}_i \quad (3)$$

Keterangan :

e = nilai *residual*

y = nilai observasi

\hat{y} = nilai prediksi

Scatter plot Normal Q-Q dapat memberikan informasi tentang tingkat linearitas masing-masing variabel bebas terhadap variabel terikat. *Scatter plot scale vs location* memiliki informasi yang sama dengan *scatter plot fitted values vs residual*, perbedaan mendasar kedua plot ini adalah nilai *residual* yang sudah distandarisasi dan memberikan *range* pada *residual*. Menurut Sinambela *et al.* (2014), rumus untuk menentukan nilai *standarized residual* adalah.

$$ZRESID = \frac{y_i - \hat{y}_i}{S_{x,y}} \quad (4)$$

Keterangan :

ZRESID = nilai *standarized residual*.

Y_i = nilai observasi.

\hat{y}_i = nilai prediksi

$S_{x,y}$ = nilai *standard error* koefisien

Scatter plot residual vs leverage memberikan informasi tentang *outlier* yang dapat mempengaruhi model secara keseluruhan dengan menghitung *Cook's Distance* masing-masing *record* yang ada pada data. O'Halloran (2010) menuliskan rumus untuk mencari *Cook's Distance* sebagai berikut.

$$D_i = \sum_{j=1}^n \frac{(\hat{y}_{j(i)} - \hat{y}_j)^2}{p\sigma^2} \quad (5)$$

Keterangan :

D_i = nilai *Cook's Distance*

$\hat{y}_{j(i)}$ = nilai mean y pada *record* j ketika observasi i dihilangkan

\hat{y}_j = nilai *mean* y pada *record* j

p = koefisien regresi

σ = nilai varian

Pengujian ini dilakukan untuk menilai apakah model yang telah dibuat dapat diimplementasi dengan data uji dan menentukan tingkat akurasi dari data. Nilai prediksi yang sudah didapatkan dari kedua skenario akan dibandingkan hasil prediksi akhirnya untuk menilai apakah model dapat memberikan hasil yang konsisten jika pembagian data uji dan data latihnya berbeda.

1.7 Tahap Evaluasi

Saat tahap pengujian telah menghasilkan nilai prediksi dan model telah memberikan atribut keterkaitan antara variabel terikat dengan variabel bebas, model akan dievaluasi tingkat *error* dari model tersebut. Hal ini dapat dilakukan dengan menghitung nilai *R-squared* dan *Root Mean Square Error* (RMSE) dari model yang dihasilkan. Rumus untuk menghitung *R-squared* menurut Hossjer (2008) adalah sebagai berikut.

$$R^2 = \frac{\sum_i (\hat{y}_i - \hat{m}_i)^2}{\sum_i (y_i - \hat{m}_i)^2} \quad (6)$$

Keterangan :

R^2 = nilai *R-squared*

\hat{y}_i = nilai prediksi

y_i = nilai aktual

\hat{m}_i = nilai rata-rata

Menurut Willmot dan Matsuura (2005), Nilai RMSE dapat dihitung dengan rumus.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (7)$$

Keterangan :

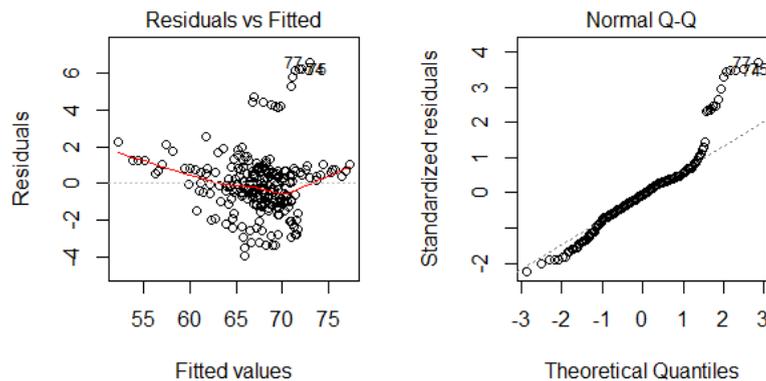
n = jumlah data.

e = nilai *residual*.

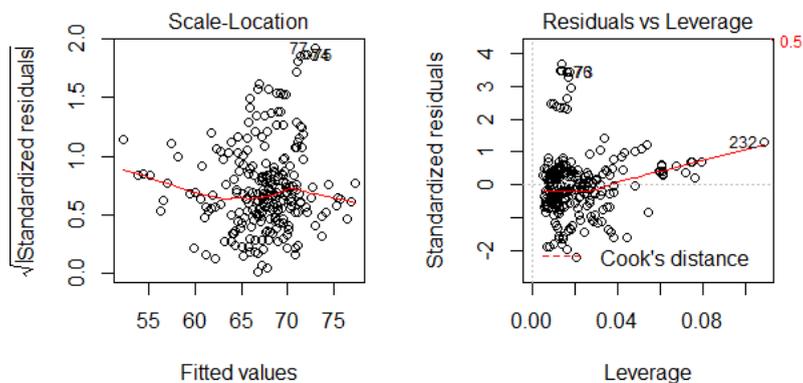
3. HASIL DAN PEMBAHASAN

1.8 Hasil Uji Regresi

Data dalam penelitian ini terdapat 272 *record* yang terbagi menjadi dua data yaitu data latih dan data uji. Variabel bebas dalam penelitian ini yaitu, harapan hidup, lama sekolah, kesehatan dan kemiskinan, sedangkan variabel terikat adalah pembangunan manusia. Untuk memenuhi skenario satu, data latih terdiri atas 238 *record* yang diambil dari tahun 2010 – 2016 dan untuk data uji terdiri atas 34 *record* tahun 2017. Visualisasi dari regresi skenario satu dapat dilihat dari gambar di bawah ini.



Gambar 2 : Scatter plot residual dan Normal Q-Q dari skenario 1



Gambar 3 : Scatter plot sebaran dan Cook's distance

Pada gambar 1, *linear residuals vs fitted* menggambarkan bahwa nilai *residual* pada masing-masing data cukup tersebar pada sumbu x dan y, serta mempunyai garis prediksi yang horizontal. Pada linier normal Q-Q, sebagian besar data berada pada garis prediksi dan memiliki beberapa *outlier* di akhir garis prediksi. Pada gambar 2, *linear scale-location* memberi informasi tentang sebaran *residual* range prediksi yang telah ditentukan oleh model dan menunjukkan nilai *residual* cukup tersebar. Linier *Residuals vs Leverage* memberi informasi tentang pengaruh *outlier* data pada hasil akhir dan menunjukkan beberapa *outlier* yang tidak terlalu mempengaruhi hasil akhir dari prediksi karena *outlier* terluar dari data masih berada dalam garis *Cook's Distance* $< 0,5$ (garis merah pada ujung kanan linier).

Dalam tahap uji regresi skenario satu didapatkan hasil uji untuk data latih seperti pada tabel berikut.

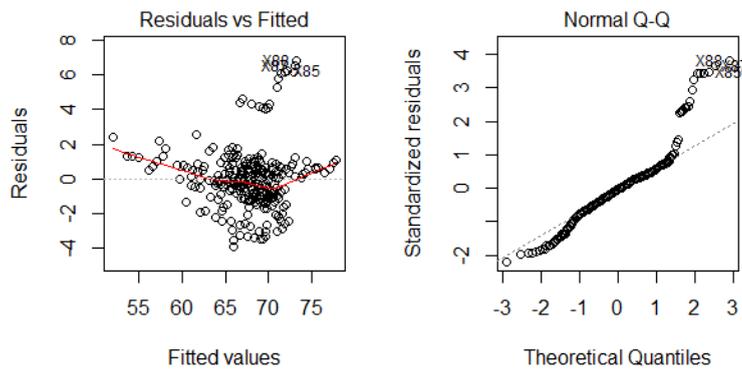
Tabel 1 : Hasil uji regresi skenario 1

Model	Estimate	Nilai Probabilitas
(Intercept)	-3.91	0.263
Harapan_Hidup (X1)	0.73	2×10^{-16}
Lama_Sekolah (X2)	1.83	2×10^{-16}
Kesehatan(X3)	0.04	0.063
Kemiskinan (X4)	-0.22	2×10^{-16}

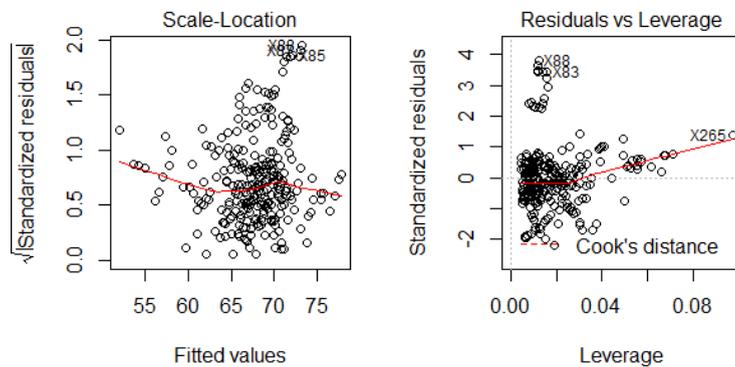
Dari hasil yang telah didapat, dapat disimpulkan model regresi dari skenario satu sebagai berikut.

$$Y_1 = 0.73X_1 + 1.83X_2 + 0.04X_3 - 0.22X_4 - 3.91 \quad (8)$$

Untuk mengolah data pada skenario dua, data dibagi menjadi data latih dan data uji melalui *k-fold cross validation* dengan nilai $k = 3$. Visualisasi skenario dua dapat dilihat pada gambar di bawah.



Gambar 4 : Scatter plot residual dan Normal Q-Q dari skenario 2



Gambar 5 : Scatter plot sebaran dan Cook's distance dari skenario 2

Pada gambar 3 menjelaskan hal yang sama dengan gambar 1 dan menunjukkan *linear residuals vs fitted* pada skenario dua memiliki sebaran yang juga cukup tersebar dan mempunyai garis prediksi yang horizontal. Linier normal Q-Q juga memiliki sebagian besar data yang berada pada garis prediksi linier dan beberapa *outlier* di akhir garis prediksi. Pada gambar 4 menjelaskan hal yang sama dengan gambar 2, linier *scale-location* juga menunjukkan nilai *residual* data cukup tersebar, serta linier *Residuals vs Leverage* juga menunjukkan ada beberapa *outlier* yang tidak terlalu mempengaruhi hasil akhir dari prediksi.

Dalam tahap uji regresi skenario dua didapatkan hasil uji untuk data latih seperti pada tabel berikut.

Tabel 2 : Hasil uji regresi

Model	Estimasi	Nilai Probabilitas
(Intercept)	-4	0.263
Harapan_Hidup (X1)	0.73	2×10^{-16}
Lama_Sekolah (X2)	1.86	2×10^{-16}
Kesehatan(X3)	0.04	0.063
Kemiskinan (X4)	-0.22	2×10^{-16}

Dari hasil yang telah dapat disimpulkan model regresi dari skenario dua sebagai berikut.

$$Y_2 = 0.73X_1 + 1.86X_2 + 0.04X_3 - 0.22X_4 - 4 \quad (9)$$

Kedua skenario di atas kemudian dilakukan uji coba dan didapatkan hasilnya dalam tabel berikut.

Tabel 3 : Hasil prediksi

Provinsi	Tahun	Data real	Skenario 1	Skenario 2
Aceh	2017	70.6	70.36	70.48
Sumatera Utara	2017	70.57	69.03	69.14
Sumatera Barat	2017	71.24	71.74	71.85
Riau	2017	71.79	71.45	71.54
Jambi	2017	69.99	70.61	70.73
Sumatera Selatan	2017	68.86	67.65	67.71
Bengkulu	2017	69.95	68.74	68.84
Lampung	2017	68.25	68.44	68.49
Bangka Belitung	2017	69.99	69.03	69.08
Kep Riau	2017	74.45	70.02	70.14
DKI Jakarta	2017	80.06	73.11	73.2
Jawa Barat	2017	70.69	71.36	71.41
Jawa Tengah	2017	70.52	72.01	72.04
DI Yogyakarta	2017	78.89	77.71	77.82
Jawa timur	2017	70.27	70.76	70.81
Banten	2017	71.42	70.36	70.45
Bali	2017	74.3	72.94	73.03
Nusa Tenggara Barat	2017	66.58	66.68	66.74

Nusa Tenggara Timur	2017	63.73	65.19	65.2
Kalimantan Barat	2017	66.26	69.52	69.61
Kalimantan Tengah	2017	69.79	69.8	69.88
Kalimantan Selatan	2017	69.65	69.13	69.17
Kalimantan Timur	2017	75.12	74.45	74.58
Kalimantan Utara	2017	69.84	72.26	72.33
Sulawesi Utara	2017	71.66	70.7	70.78
Sulawesi Tengah	2017	68.11	67.46	67.51
Sulawesi Selatan	2017	70.34	70.56	70.67
Sulawesi Tenggara	2017	69.86	70.49	70.6
Gorontalo	2017	67.01	66.77	66.79
Sulawesi Barat	2017	64.3	64.72	64.79
Maluku	2017	68.19	66.27	66.41
Maluku Utara	2017	67.2	69.67	69.86
Papua Barat	2017	62.99	62.39	62.46
Papua	2017	59.09	57.76	57.79

1.9 Evaluasi Model

Evaluasi model yang dilakukan dalam penelitian ini ditentukan dengan menghitung nilai *R-Squared* dan *Root Mean Square Error* (RMSE) pada masing-masing model. Hasil dari nilai tersebut dapat dilihat pada tabel berikut.

Tabel 4 : Hasil Evaluasi

Model	Skenario 1	Skenario 2
<i>R-Squared</i>	0.84	0.83
RMSE	1.84	1.82

R-Squared menyatakan tingkat akurasi dari model regresi yang sudah didapatkan dan menunjukkan hasil akurasi sebesar 84% untuk skenario satu dan 83% untuk skenario dua. RMSE menyatakan nilai rata-rata besarnya kesalahan yang dihasilkan oleh suatu model prakira dan menunjukkan hasil RMSE sebesar 1.84 untuk skenario satu dan 1.82 untuk skenario dua.

4. KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan dapat diketahui bahwa variabel bebas yang paling mempengaruhi nilai variabel terikat adalah X2 atau lama sekolah karena memiliki nilai probabilitas paling mendekati nol dan nilai *estimate* terbesar dan evaluasi model yang digunakan dalam penelitian ini menghasilkan tingkat akurasi dengan *R-Squared* sebesar 0.84 (84%) dan *Root Mean Square Error* (RMSE) sebesar 1.84 pada skenario satu, sedangkan skenario dua memiliki nilai *R-Squared* sebesar 0.83 (83%) dan RMSE sebesar 1.82.

Berdasarkan penelitian yang sudah dilakukan maka saran yang diharapkan adalah penelitian dapat dikembangkan dengan diperluas obyek penelitiannya seperti menambahkan variabel dan menggunakan metode lain dalam memprediksi nilai IPM.

Referensi

Amrin. 2016. "Data Mining dengan Regresi Linier Berganda untuk Peramalan Tingkat Inflasi" dalam jurnal Techno Nusa Mandiri Vol. XIII No. 1

- Badan Pusat Statistik. 2018 . <http://www.bps.go.id>. Diakses: 30 Oktober 2018.
- Bardja, Sutiati. 2017. "Pengaruh Penerapan Senam *Hook Ups* Terhadap Tingkat Percaya Diri Anak Kelas Dua Min Guwa Kidul" dalam jurnal Ilmiah Indonesia Vol.2 No. 12.
- Darsyah, Moh. Yamin. 2017. "Klasifikasi Indeks Pembangunan Manusia (IPM) dengan Pendekatan *K-Nearset Neighbour* (K-NN). Seminar Nasional Pendidikan, Sains, dan Teknologi : 29 – 35.
- Hossjer, Ola. 2008. "*On the coefficient of determination for mixed regression models*" dalam jurnal *Statistical Planning and Inference* Vol. 138.
- O'Halloran, Sharyn. 2010. "*Model Checking*" [Dokumen PowerPoint]. diambil dari <http://www.columbia.edu>.
- Putra, Dwi Maunere, Vita Ratnasari. 2015. "Pemodelan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur dengan menggunakan Metode Regresi Logistik Ridge" dalam jurnal Sains dan Seni ITS Vol.4 No. 2.
- Sari, Betha Nurina, Priati. 2016."Identifikasi Keterkaitan Variabel dan Prediksi Indeks Pembangunan Manusia (IPM) Provinsi Jawa Barat Menggunakan *Dynamic Bayesian Networks*" dalam jurnal Infotel Vol.8 No.2.
- Sinambela, Sabam Daoni, Suwarno Ariswoyo. 2014. "Menentukan Koefisien Determinasi Antara Estimasi M dengan *Type Welsch* Dengan *Least Trimmed Square* dalam Data yang Mempunyai Pencilan" dalam jurnal Sainia Matematika Vol. 02 No. 03.
- Sugiyono. (2014). Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta.
- Syilfi, Dwi Ispriyanti, Diah Safitri. 2012. "Analisis Regresi Linier *Piecewise* Dua Segmen" dalam jurnal Gaussian Vol. 1 No. 1.
- United Nations Development Programme (UNDP). 1990. *Global Human Development Report. Human Resources Department*.
- Willmott, C. and Matsuura, K. 2005. "*Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance*" dalam jurnal Climate Research Vol.30.